WORLDWIDE wwPDB PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report ⓘ

Nov 2, 2024 – 11:56 AM EDT

PDB ID	:	6BOZ
Title	:	Structure of human SETD8 in complex with covalent inhibitor MS4138
Authors	:	Babault, N.; Anqi, M.; Jin, J.
Deposited on	:	2017-11-21
Resolution	:	2.40 Å(reported)

The following versions of software and data (see references ⓘ) were used in the production of this report:

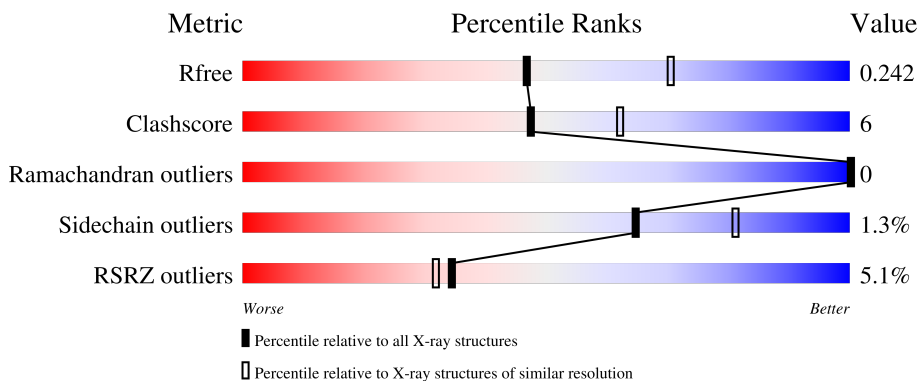|		
MolProbity	:	4.02b-467
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	1.20.1
EDS	:	3.0
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.003 (Gargrove)
Density-Fitness	:	1.0.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.39

1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.40 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.

Metric	Percentile Ranks	Value
Rfree		0.242
Clashscore		6
Ramachandran outliers		0
Sidechain outliers		1.3%
RSRZ outliers		5.1%

Worse   *Better*

▌ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	164625	4642 (2.40-2.40)
Clashscore	180529	5218 (2.40-2.40)
Ramachandran outliers	177936	5158 (2.40-2.40)
Sidechain outliers	177891	5159 (2.40-2.40)
RSRZ outliers	164620	4642 (2.40-2.40)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments of the lower bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria respectively. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions <=5% The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	189	3% / 63% 9% • 26%
1	B	189	4% / 61% 10% 29%

2 Entry composition ⓘ

There are 4 unique types of molecules in this entry. The entry contains 2299 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called N-lysine methyltransferase KMT5A.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	139	Total	C	N	O	S	0	0	0
			1099	686	191	217	5			
1	B	134	Total	C	N	O	S	0	0	0
			1062	665	182	210	5			

There are 56 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	205	MET	-	initiating methionine	UNP Q9NQR1
A	206	SER	-	expression tag	UNP Q9NQR1
A	207	TYR	-	expression tag	UNP Q9NQR1
A	208	TYR	-	expression tag	UNP Q9NQR1
A	209	HIS	-	expression tag	UNP Q9NQR1
A	210	HIS	-	expression tag	UNP Q9NQR1
A	211	HIS	-	expression tag	UNP Q9NQR1
A	212	HIS	-	expression tag	UNP Q9NQR1
A	213	HIS	-	expression tag	UNP Q9NQR1
A	214	HIS	-	expression tag	UNP Q9NQR1
A	215	ASP	-	expression tag	UNP Q9NQR1
A	216	TYR	-	expression tag	UNP Q9NQR1
A	217	ASP	-	expression tag	UNP Q9NQR1
A	218	ILE	-	expression tag	UNP Q9NQR1
A	219	PRO	-	expression tag	UNP Q9NQR1
A	220	THR	-	expression tag	UNP Q9NQR1
A	221	THR	-	expression tag	UNP Q9NQR1
A	222	GLU	-	expression tag	UNP Q9NQR1
A	223	ASN	-	expression tag	UNP Q9NQR1
A	224	LEU	-	expression tag	UNP Q9NQR1
A	225	TYR	-	expression tag	UNP Q9NQR1
A	226	PHE	-	expression tag	UNP Q9NQR1
A	227	GLN	-	expression tag	UNP Q9NQR1
A	228	GLY	-	expression tag	UNP Q9NQR1
A	229	ALA	-	expression tag	UNP Q9NQR1

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	230	MET	-	expression tag	UNP Q9NQR1
A	231	GLY	-	expression tag	UNP Q9NQR1
A	343	SER	CYS	engineered mutation	UNP Q9NQR1
B	205	MET	-	initiating methionine	UNP Q9NQR1
B	206	SER	-	expression tag	UNP Q9NQR1
B	207	TYR	-	expression tag	UNP Q9NQR1
B	208	TYR	-	expression tag	UNP Q9NQR1
B	209	HIS	-	expression tag	UNP Q9NQR1
B	210	HIS	-	expression tag	UNP Q9NQR1
B	211	HIS	-	expression tag	UNP Q9NQR1
B	212	HIS	-	expression tag	UNP Q9NQR1
B	213	HIS	-	expression tag	UNP Q9NQR1
B	214	HIS	-	expression tag	UNP Q9NQR1
B	215	ASP	-	expression tag	UNP Q9NQR1
B	216	TYR	-	expression tag	UNP Q9NQR1
B	217	ASP	-	expression tag	UNP Q9NQR1
B	218	ILE	-	expression tag	UNP Q9NQR1
B	219	PRO	-	expression tag	UNP Q9NQR1
B	220	THR	-	expression tag	UNP Q9NQR1
B	221	THR	-	expression tag	UNP Q9NQR1
B	222	GLU	-	expression tag	UNP Q9NQR1
B	223	ASN	-	expression tag	UNP Q9NQR1
B	224	LEU	-	expression tag	UNP Q9NQR1
B	225	TYR	-	expression tag	UNP Q9NQR1
B	226	PHE	-	expression tag	UNP Q9NQR1
B	227	GLN	-	expression tag	UNP Q9NQR1
B	228	GLY	-	expression tag	UNP Q9NQR1
B	229	ALA	-	expression tag	UNP Q9NQR1
B	230	MET	-	expression tag	UNP Q9NQR1
B	231	GLY	-	expression tag	UNP Q9NQR1
B	343	SER	CYS	engineered mutation	UNP Q9NQR1

- Molecule 2 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: $C_2H_6O_2$).

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	A	1	Total C O			0	0
			4	2	2		

- Molecule 3 is N-(3-{[7-(2-aminoethoxy)-6-methoxy-2-(pyrrolidin-1-yl)quinazolin-4-yl]amino}propyl)prop-2-enamide (three-letter code: E1J) (formula: $C_{21}H_{30}N_6O_3$).
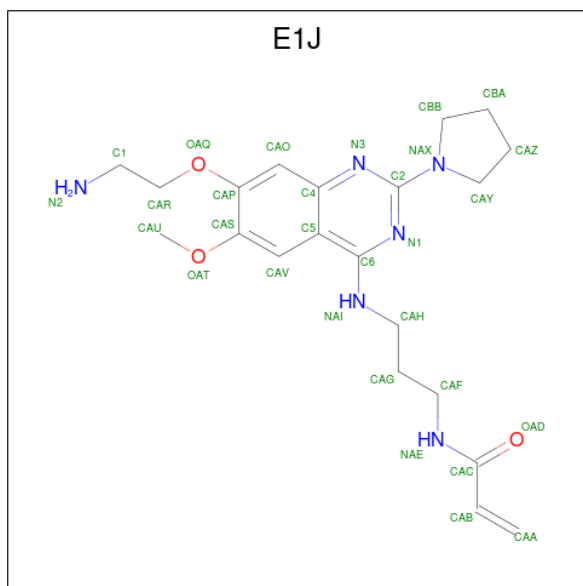
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	A	1	Total	C	N	O	0	0
			30	21	6	3		
3	B	1	Total	C	N	O	0	0
			30	21	6	3		

- Molecule 4 is water.

| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	45	Total O 45 45		0	0
4	B	29	Total O 29 29		0	0

3 Residue-property plots ⓘ

These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density (RSRZ > 2). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

• Molecule 1: N-lysine methyltransferase KMT5A
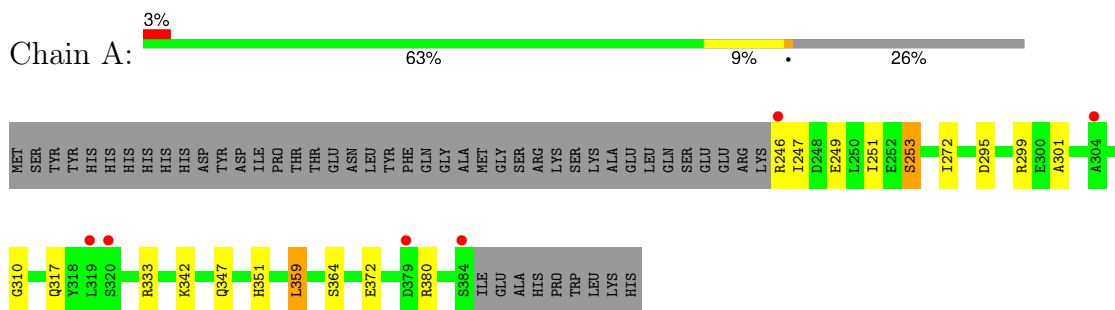
• Molecule 1: N-lysine methyltransferase KMT5A
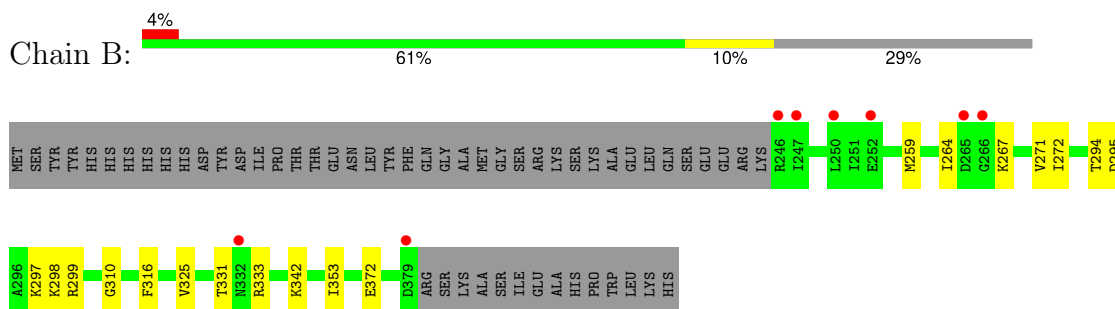
4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	31.56Å 68.06Å 125.90Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	62.95 – 2.40 62.95 – 2.40	Depositor EDS
% Data completeness (in resolution range)	99.5 (62.95-2.40) 99.5 (62.95-2.40)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.11	Depositor
$< I/\sigma(I) >$ [1]	2.98 (at 2.40Å)	Xtriage
Refinement program	PHENIX 1.11.1_2575	Depositor
R, R_{free}	0.179 , 0.242 0.179 , 0.242	Depositor DCC
R_{free} test set	579 reflections (5.15%)	wwPDB-VP
Wilson B-factor (Å^2)	30.0	Xtriage
Anisotropy	0.237	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.35 , 51.9	EDS
L-test for twinning[2]	$< |L| > = 0.49, < L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	2299	wwPDB-VP
Average B, all atoms (Å^2)	38.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 9.09% of the height of the origin peak. No significant pseudotranslation is detected.*

---

[1]Intensities estimated from amplitudes.
[2]Theoretical values of $< |L| >$, $< L^2 >$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5   Model quality ⓘ

5.1   Standard geometry ⓘ

Bond lengths and bond angles in the following residue types are not validated in this section: E1J, EDO

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	#$|Z|$ >5	RMSZ	#$|Z|$ >5
1	A	0.40	0/1116	0.58	1/1499 (0.1%)
1	B	0.38	0/1079	0.56	0/1451
All	All	0.39	0/2195	0.57	1/2950 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	359	LEU	CA-CB-CG	-5.05	103.67	115.30

There are no chirality outliers.

There are no planarity outliers.

5.2   Too-close contacts ⓘ

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry-related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	1099	0	1077	11	0
1	B	1062	0	1036	15	0
2	A	4	0	6	0	0
3	A	30	0	0	0	0
3	B	30	0	0	0	0
4	A	45	0	0	2	0
4	B	29	0	0	2	0
All	All	2299	0	2119	24	0

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 6.

The worst 5 of 24 close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:264:ILE:HB	1:B:267:LYS:HB3	1.63	0.79
1:B:272:ILE:HD13	1:B:372:GLU:HA	1.83	0.60
1:A:347:GLN:HG3	1:A:364:SER:HB3	1.84	0.59
1:A:301:ALA:O	1:B:342:LYS:NZ	2.36	0.59
1:A:333:ARG:NH2	4:A:502:HOH:O	2.36	0.58

There are no symmetry-related clashes.

5.3 Torsion angles ⓘ

5.3.1 Protein backbone ⓘ

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	A	137/189 (72%)	130 (95%)	7 (5%)	0	100	100
1	B	132/189 (70%)	125 (95%)	7 (5%)	0	100	100
All	All	269/378 (71%)	255 (95%)	14 (5%)	0	100	100

There are no Ramachandran outliers to report.

5.3.2 Protein sidechains ⓘ

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	A	117/162 (72%)	114 (97%)	3 (3%)	41	62
1	B	113/162 (70%)	113 (100%)	0	100	100
All	All	230/324 (71%)	227 (99%)	3 (1%)	65	81

All (3) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	253	SER
1	A	317	GLN
1	A	380	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	B	305	GLN

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates ⓘ

There are no oligosaccharides in this entry.

5.6 Ligand geometry ⓘ

3 ligands are modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	#|Z| > 2	Counts	RMSZ	#|Z| > 2
3	E1J	A	402	1	32,32,32	1.43	5 (15%)	38,42,42	2.70	13 (34%)
2	EDO	A	401	-	3,3,3	0.54	0	2,2,2	0.37	0
3	E1J	B	401	1	32,32,32	1.48	6 (18%)	38,42,42	2.65	13 (34%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
3	E1J	A	402	1	-	2/20/27/27	0/3/3/3
2	EDO	A	401	-	-	0/1/1/1	-
3	E1J	B	401	1	-	2/20/27/27	0/3/3/3

The worst 5 of 11 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	B	401	E1J	CAA-CAB	4.66	1.52	1.30
3	A	402	E1J	CAA-CAB	4.26	1.50	1.30
3	A	402	E1J	C6-C5	-3.58	1.40	1.44
3	B	401	E1J	C6-C5	-3.36	1.40	1.44
3	A	402	E1J	CAB-CAC	2.81	1.52	1.48

The worst 5 of 26 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	A	402	E1J	CAA-CAB-CAC	-10.55	110.63	122.22
3	B	401	E1J	CAA-CAB-CAC	-10.35	110.85	122.22
3	A	402	E1J	C6-C5-C4	5.27	118.97	115.86
3	A	402	E1J	CAV-C5-C6	-4.25	121.38	124.84
3	B	401	E1J	C6-C5-C4	4.05	118.25	115.86

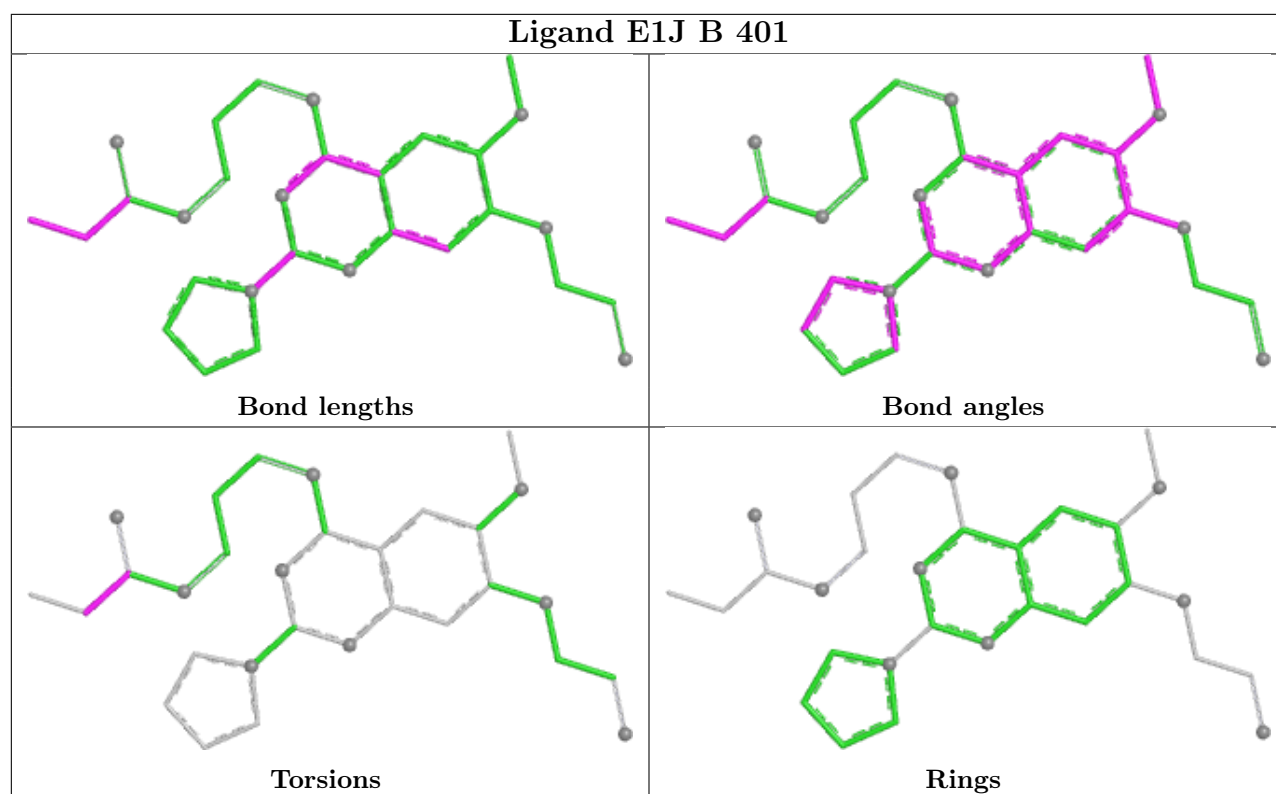There are no chirality outliers.

All (4) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	402	E1J	CAA-CAB-CAC-NAE
3	A	402	E1J	CAA-CAB-CAC-OAD
3	B	401	E1J	CAA-CAB-CAC-NAE
3	B	401	E1J	CAA-CAB-CAC-OAD

There are no ring outliers.

No monomer is involved in short contacts.

The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included. For torsion angles, if less then 5% of the Mogul distribution of torsion angles is within 10 degrees of the torsion angle in question, then that torsion angle is considered an outlier. Any bond that is central to one or more torsion angles identified as an outlier by Mogul will be highlighted in the graph. For rings, the root-mean-square deviation (RMSD) between the ring in question and similar rings identified by Mogul is calculated over all ring torsion angles. If the average RMSD is greater than 60 degrees and the minimal RMSD between the ring in question and any Mogul-identified rings is also greater than 60 degrees, then that ring is considered an outlier. The outliers are highlighted in purple. The color gray indicates Mogul did not find sufficient equivalents in the CSD to analyse the geometry.

Ligand E1J A 402

Bond lengths

Bond angles

Torsions

Rings

Ligand E1J B 401

Bond lengths   Bond angles

Torsions   Rings
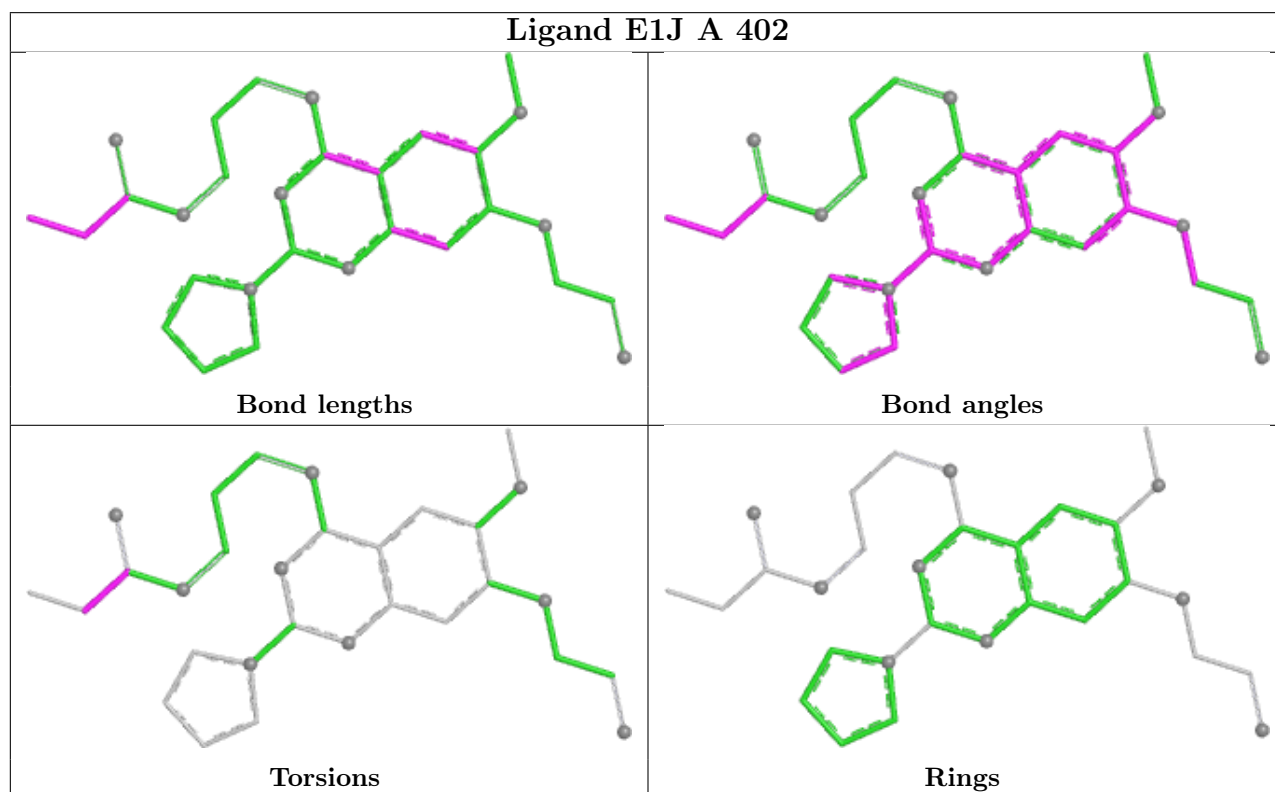
5.7   Other polymers ⓘ

There are no such residues in this entry.

5.8   Polymer linkage issues ⓘ

There are no chain breaks in this entry.

6   Fit of model and data ⓘ

6.1   Protein, DNA and RNA chains ⓘ

In the following table, the column labelled '#RSRZ> 2' contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95th percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled 'Q< 0.9' lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2			OWAB(Å²)	Q<0.9
1	A	139/189 (73%)	-0.16	6 (4%)	40	38	18, 33, 78, 102	0
1	B	134/189 (70%)	-0.12	8 (5%)	29	26	18, 34, 76, 90	0
All	All	273/378 (72%)	-0.14	14 (5%)	34	32	18, 33, 77, 102	0

The worst 5 of 14 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	384	SER	3.7
1	A	304	ALA	3.1
1	B	247	ILE	3.0
1	A	319	LEU	2.5
1	B	332	ASN	2.5

6.2   Non-standard residues in protein, DNA, RNA chains ⓘ

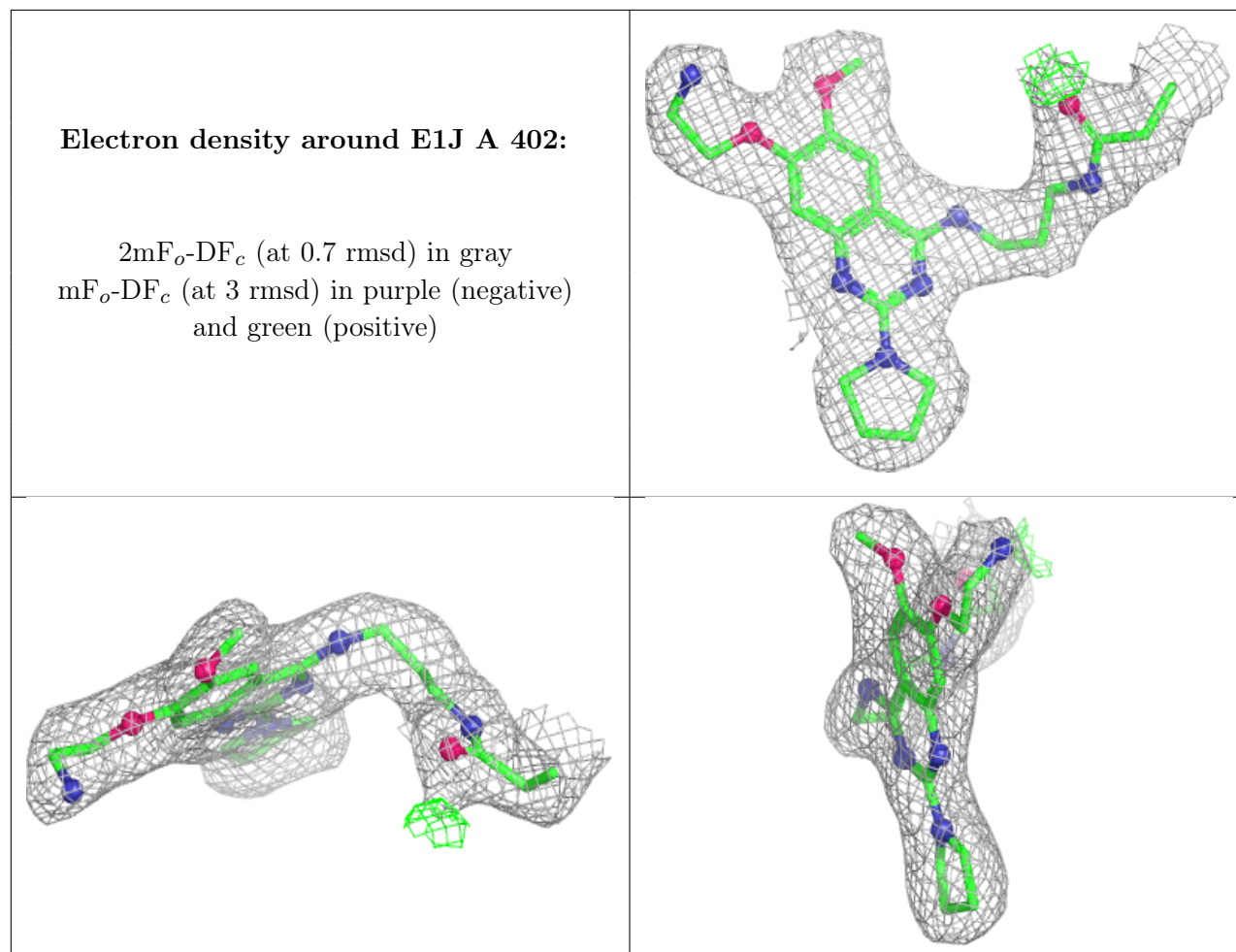There are no non-standard protein/DNA/RNA residues in this entry.

6.3   Carbohydrates ⓘ

There are no monosaccharides in this entry.
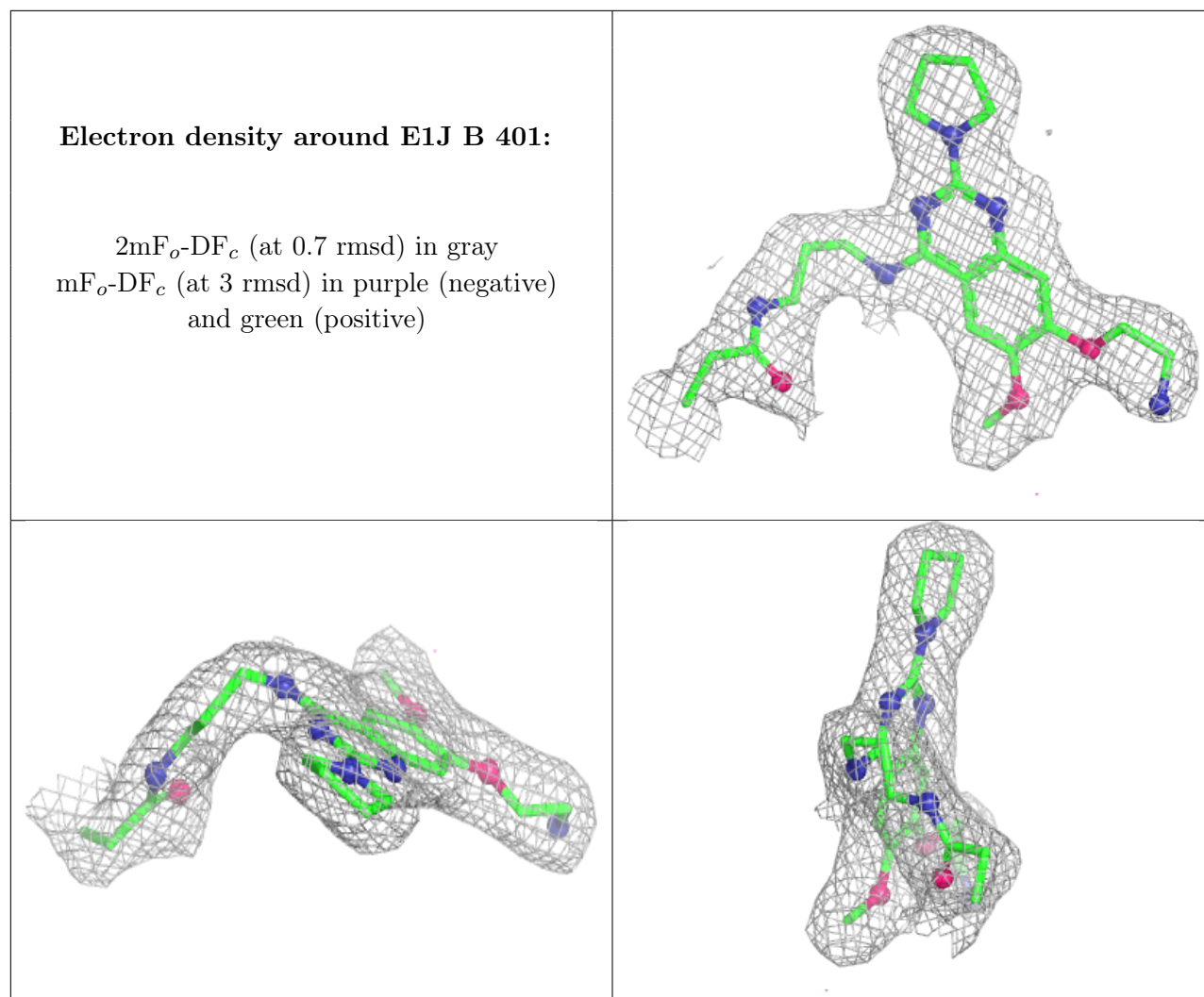
6.4   Ligands ⓘ

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors($Å^2$)	Q<0.9
2	EDO	A	401	4/4	0.96	0.06	13,16,17,20	0
3	E1J	A	402	30/30	0.97	0.05	11,19,23,25	0
3	E1J	B	401	30/30	0.97	0.06	9,21,32,37	0

The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation to approximate a three-dimensional view.

Electron density around E1J A 402:

$2mF_o\text{-}DF_c$ (at 0.7 rmsd) in gray
$mF_o\text{-}DF_c$ (at 3 rmsd) in purple (negative)
and green (positive)

Electron density around E1J B 401:

2mF$_o$-DF$_c$ (at 0.7 rmsd) in gray
mF$_o$-DF$_c$ (at 3 rmsd) in purple (negative)
and green (positive)

6.5 Other polymers ⓘ

There are no such residues in this entry.